

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada tahun 1990 hingga 2013, jumlah angka kematian yang disebabkan oleh penyakit kardiovaskular naik sebesar 41%, dari 12,3 juta sampai 17,3 juta jiwa didunia. Angka kematian yang besar ini, 50% kasusnya terjadi di Amerika Serikat dan negara berkembang lainnya, serta terus meningkat dari tahun ke tahun [1]. Penyakit kardiovaskular merupakan penyakit yang menyerang jantung dan pembuluh darah, sehingga dapat menyebabkan nyeri pada dada atau bahkan serangan jantung [2]. Padahal jantung merupakan organ yang berfungsi untuk memompa darah ke seluruh jaringan tubuh. Jika kinerja jantung terganggu atau tidak bekerja dengan baik, maka organ tubuh vital berupa otak dan ginjal tentu akan mengalami gangguan yang serius. Berhentinya organ jantung dapat menyebabkan manusia mengalami kematian dalam beberapa menit. Penyakit jantung tidak menular seperti halnya flu atau demam berdarah, namun penyakit ini sering terjadi pada orang dewasa atau lanjut usia, bahkan kini, orang yang masih usia muda rentan terkena penyakit jantung.

Dokter menyebutkan ada beberapa faktor yang beresiko menyebabkan penyakit jantung diantaranya tekanan darah tinggi, kelebihan berat badan, merokok, serta jarang berolahraga. Namun kebanyakan orang cenderung tidak menyadari secara dini, gejala orang yang terkena penyakit kardiovaskular sampai mereka merasakan nyeri pada dada, memiliki tekanan darah tinggi atau bahkan serangan jantung, yang biasanya baru mereka rasakan pada usia lanjut. Oleh karena itu, orang-orang harus menyadari pentingnya tindakan dini untuk melihat faktor-faktor apa saja yang dapat menyebabkan penyakit jantung [3].

Dalam bidang kedokteran, para ahli tenaga medis tentu memiliki informasi tentang gejala atau karakteristik dari penyakit jantung. Kegiatan yang mengumpulkan atau memakai data untuk menemukan keteraturan, pola atau hubungan dalam set data untuk mendapatkan sebuah informasi baru disebut sebagai *data mining*. Selain itu, data mining juga berfungsi mengklasifikasikan data untuk menentukan keputusan dari informasi baru yang sudah didapatkan dari pengolahan data menggunakan perhitungan algoritma. Teknik klasifikasi ini merupakan teknik

yang memetakan data teks atau gambar, kedalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya sehingga dapat mudah dianalisis dan didalamnya terdapat satu atribut tujuan atau disebut sebagai atribut label. Dalam data baru (*data testing*), atribut label akan dicari dengan berdasarkan atribut lain pada data lama (*data training*) [4].

Beberapa penelitian terkait identifikasi data telah dilakukan dalam beberapa tahun sebelumnya. Penelitian tahun 2015, penggunaan metode *K-nearest Neighbour* dan *Support Vector Machine* untuk mengklasifikasikan penyakit *Diabetes Retinopati* berdasarkan *citra eksudat* dan *microneurisma*. Perbandingan kedua metode tersebut menunjukkan bahwa *K-Nearest Neighbour* lebih unggul dari pada *Support Vector Machine*[5]. Sedangkan untuk penyakit jantung, beberapa penelitian pernah dilakukan menggunakan metode *logistic regression*[6], *Neural network* [6], *Deep learning*[7], *Decision Tree*[8] dan *Artificial neural Network* [8].

Untuk klasifikasi penyakit *Diabetes Mellitus* menggunakan metode *K-Nearest Neighbour* dan *Gain Ration* pada tahun 2017, yang dimana hasil akurasi tertingginya mencapai 75,26% [4]. Metode *K-Nearest Neighbour* juga digunakan untuk klasifikasi pada bunga[9], kelulusan siswa[10], dan penerima beasiswa yang nilai akurasinya cukup tinggi[11].

Beberapa penelitian telah dilakukan untuk mendapatkan hasil maksimal dalam mengidentifikasi penyakit jantung. Namun berbagai metode telah mendapatkan hasil yang beragam. Diantara metode identifikasi yang cukup populer untuk digunakan yakni *K-Nearest Neighbour*. Metode ini menghitung kedekatan antar data baru dengan data lama berdasarkan pencocokan bobot sehingga metode *K-Nearest Neighbour* memiliki konsistensi yang kuat serta efektif dalam mengolah data yang besar dan memberikan hasil yang akurat [11].

Dataset penyakit jantung memiliki tipe atribut yang beragam, yakni data numerik, data kategorik dan data biner. Untuk mengatasi masalah tersebut, data perlu dinormalisasikan untuk mendapatkan nilai yang lebih mudah untuk di proses. Kemudian untuk data numerik, dapat dihitung menggunakan *Euclidean Distance*, untuk data biner, dapat dihitung menggunakan *Jaccard Distance*, sedangkan untuk data kategorik, dapat dihitung dengan *Hamming Distance*. Dimana ketiga metode perhitungan jarak ketetanggaan atau kedekatan tersebut, dapat diimplementasikan

kedalam metode *K-Nearest Neighbour*. Oleh karena itu, penelitian ini akan mengidentifikasi data penyakit jantung dengan menggunakan metode *K-Nearest Neighbour*.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Permasalahan yang terjadi dalam penelitian ini adalah :

1. Dataset penyakit jantung memiliki tipe atribut yang beragam.
2. Dibutuhkan tingkat akurasi yang baik untuk identifikasi penyakit jantung.

1.2.2 Metode Usulan

Metode yang diusulkan untuk mengidentifikasi data penyakit jantung dalam penelitian ini adalah *K-Nearest Neighbour*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan metode yang diusulkan, maka didapatkan pertanyaan bahwa seberapa tinggi tingkat akurasi yang didapatkan dengan menggunakan metode *K-Nearest Neighbour* untuk mengidentifikasi penyakit jantung?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Tujuan penelitian ini adalah mengidentifikasi penyakit jantung dari dataset yang didapatkan dengan menggunakan metode *K-Nearest Neighbour*.

1.3.2 Manfaat Penelitian

Dari penelitian ini, diharapkan mampu memberikan manfaat dalam membangun ataupun mengembangkan sistem yang dapat mengidentifikasi penyakit jantung dengan tingkat akurasi yang baik, sehingga dapat menentukan tindakan dini untuk pasien penyakit jantung.

1.4 Batasan – Batasan

Batasan masalah penelitian ini yaitu :

1. *Dataset* penyakit jantung memiliki 1025 sampel dan diambil dari <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>.
2. *Dataset* memiliki tipe atribut yang beragam, yakni data numerik, data kategori dan data biner dalam satu *dataset*.
3. Sistem dibangun menggunakan tools **Orange data mining 3.2**.

1.5 Sistematika Proposal

Sistematika penulisan proposal ini terdiri dari beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang mengenai ruang lingkup masalah yang dibahas, solusi yang diberikan untuk dapat menyelesaikan permasalahan, tujuan dan manfaat dengan diterapkannya, solusi yang diusulkan, batasan solusi yang diberikan, serta sistematika penulisan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori mengenai metode yang digunakan dan penelitian – penelitian terkait yang telah dilakukan sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan yakni *K-Nearest Neighbours (KNN)*. selain itu, bab III juga menjelaskan arsitektur dari metode, dataset yang digunakan, perhitungan dan uji coba manual menggunakan *Microsoft Excel* serta tahapan – tahapan yang dilakukan dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang bagaimana sistem dibuat, hasil pengujian berdasarkan skenario pengujian dan analisa.

BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi tentang kesimpulan dari analisa yang telah dilakukan dan saran untuk penelitian selanjutnya dalam mengidentifikasi penyakit jantung

REFERENSI

Berisi tentang daftar jurnal atau hasil penelitian sebelumnya yang dijadikan acuan dalam penelitian ini.